CONFIDENCE BANDS FOR THE MEDIAN SURVIVAL TIME AS A FUNCTION OF THE COVARIATES IN THE COX MODEL

BY

DEBORAH BURR and HANI DOSS

TECHNICAL REPORT NO. 443
JUNE 25, 1991

PREPARED UNDER CONTRACT
N00014-89-J-1627 (NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

CONFIDENCE BANDS FOR THE MEDIAN SURVIVAL TIME AS A FUNCTION OF THE COVARIATES IN THE COX MODEL

BY

DEBORAH BURR and HANI DOSS

TECHNICAL REPORT NO. 443
JUNE 25, 1991

Prepared Under Contract
N00014-89-J-1627 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

1 Introduction and Summary

The proportional hazards model of Cox (1972) specifies that the hazard rate for an individual with covariate vector x is

$$\lambda(t|x) = \lambda(t)\exp(\beta_0'x) \tag{1.1}$$

where β_0 is a vector of unknown regression coefficients and λ, the underlying baseline hazard rate, is an unknown and unspecified nonnegative function. In most cases where this model is used, interest centers on the estimation of β_0. However, it is often also very useful to investigate the effect of the covariates on the median (or some other fixed quantile) of the survival time. Let $\xi_p(x)$ be the p^{th} quantile of the distribution of the lifelength of an individual with covariate vector x.

A first attempt to estimate $\xi_p(x)$ proceeds as follows. Let $\Lambda(t|x)$ and $S(t|x)$ be the cumulative hazard function and the survival function, respectively, associated with $\lambda(t|x)$, and let $\Lambda(t)$ be the cumulative hazard function associated with λ. We then have

$$\Lambda(t|x) = \Lambda(t)\exp(\beta_0'x). \tag{1.2}$$

Using the continuity implied by (1.1) we may write

$$S(t|x) = \exp(-\Lambda(t)\exp(\beta_0'x)), \tag{1.3}$$

and solving the equation $S(t|x) = 1 - p$ we obtain

$$\xi_p(x) = \Lambda^{-1}([-\log(1-p)]\exp(-\beta_0'x)). \tag{1.4}$$

The natural estimate for the right side of (1.4) is

$$\bar{\xi}_p(x) = \hat{\Lambda}^{-1}([-\log(1-p)]\exp(-\hat{\beta}'x)), \tag{1.5}$$

where $\hat{\beta}$ is Cox's (1972) maximum partial likelihood estimate of β_0 and $\hat{\Lambda}$ is the usual "Nelson-Aalen" type estimator of Λ (see equation (2.4) of Section 2). (In (1.4), (1.5), and throughout the paper, for an arbitrary increasing function f, f^{-1} denotes the right continuous inverse of f defined by $f^{-1}(t) = \sup\{s : f(s) \le t\}$). See Cox and Oakes (1984, pp. 108), Miller and Halpern (1983), and Dabrowska and Doksum (1987). See also Tsiatis (1981) for the related problem of estimating $S(t|x)$.

A second approach is to note that for an arbitrary cumulative hazard function H, the survival function corresponding to H is the product integral

$$S(t) = \prod_{s \le t}(1 - H(ds)), \tag{1.6}$$

(see Gill and Johansen (1989) or Kalbfleisch and Prentice (1980, sec. 1.2.3)) and that the p^{th} quantile of $1 - S$ is $(1-S)^{-1}(p)$. For the case of the hazard function $\Lambda(t|x)$ given by (1.2), this gives

$$\xi_p(x) = \sup\Big\{s : 1 - \prod_{u \le s}(1 - \Lambda(du)\exp(\beta_0'x)) \le p\Big\}. \tag{1.7}$$

Substituting $\hat{\Lambda}$ for Λ and $\hat{\beta}$ for β_0, we obtain

$$\tilde{\xi}_p(x) = \sup\Big\{s : 1 - \prod_{u \leq s}(1 - \hat{\Lambda}(du)\exp(\hat{\beta}'x)) \leq p\Big\}. \tag{1.8}$$

The estimate (1.8) has sounder theoretical basis than does (1.5). Let $\mathcal{H}$ be the space of all cumulative hazard functions and $\mathcal{R}$ be the real line. The function mapping an arbitrary point $(\Lambda, \beta) \in \mathcal{H} \times \mathcal{R}$ into $\xi_p(x)$ is (1.7) *not* (1.4), since (1.4) is valid only if Λ is continuous.

Before settling on $\tilde{\xi}_p(x)$ there is a slightly subtle point that needs to be raised. The estimator $\tilde{\xi}_p(x)$ was obtained under the assumption that the Cox model is given by (1.2). However, there is another way to specify the model, namely

$$S(t|x) = S(t)^{\exp(\beta_0'x)}, \tag{1.9}$$

where $S(t)$ is a baseline survival function. In general, the models (1.2) and (1.9) are different, although they agree if S (or Λ) is continuous. Because the estimates of Λ and of β_0 give rise to discrete distributions, it is important to decide on the appropriate specification of the Cox model in general. We note that if continuous survival times $T_1, \ldots, T_n$ arising from the Cox model are recorded in a discrete way (e.g. the survival times are recorded in days), then the resultant survival times do not follow (1.2) but do follow (1.9) (see Kalbfleisch and Prentice (1980, sec. 4.6)). So we take (1.9) as the specification of the Cox model in general. An elementary calculation yields that in this setup, the cumulative hazard for an individual with covariate x satisfies

$$1 - \Lambda(dt|x) = (1 - \Lambda(dt))^{\exp(\beta_0'x)} \tag{1.10}$$

where Λ corresponds to the S of (1.9) via (1.6), or equivalently

$$\Lambda(t) = -\int_{[0,t]} \frac{dS(s)}{S(s-)}. \tag{1.11}$$

The estimate

$$\hat{\xi}_p(x) = \sup\Big\{s : 1 - \prod_{u \leq s}(1 - \hat{\Lambda}(du))^{\exp(\hat{\beta}'x)} \leq p\Big\} \tag{1.12}$$

is obtained by substituting $\hat{\Lambda}$ for Λ and $\hat{\beta}$ for β_0 (other estimates of Λ can be used; see for example Kalbfleisch and Prentice (1980, sec. 4.3). Our Monte Carlo studies reported in Section 4 show that in terms of mean squared error, $\hat{\xi}_p(x)$ is slightly better than $\tilde{\xi}_p(x)$ and that each of these noticeably outperforms $\bar{\xi}_p(x)$. The three estimates $\bar{\xi}_p(x)$, $\tilde{\xi}_p(x)$, and $\hat{\xi}_p(x)$ all have the same first order asymptotics.

Dabrowska and Doksum (1987) considered estimation of $\xi_p(x)$. They studied the estimator $\bar{\xi}_p(x)$ (more precisely, the estimate (1.5) with Breslow's (1972, 1974) estimate of Λ instead of $\hat{\Lambda}$) and obtained the asymptotic distribution of $\sqrt{n}(\bar{\xi}_p(x) - \xi_p(x))$ as a process in p, from which they constructed asymptotic confidence intervals for $\xi_p(x)$. They also provide information on the efficiency of the semiparametric estimator $\bar{\xi}_p(x)$ vs. the optimal estimator in certain parametric models.

In this paper we expand on the work of Dabrowska and Doksum, and develop confidence bands for the function $\xi_p(x)$ as x varies. We propose two types of confidence bands (more properly, confidence sets). One type we refer to as "simulated-process bands"; the other is based on the bootstrap. The development of the simulated-process bands is based on weak convergence results of the following type. Let p be fixed and let x vary over a set K. Then, as $n \to \infty$,

$$\sqrt{n}(\hat{\xi}_p(x) - \xi_p(x)) \longrightarrow V_p(x) \quad \text{in distribution} \tag{1.13}$$

where V_p is a Gaussian process defined on K. Since x is multidimensional, this result involves weak convergence of multiparameter stochastic processes. (A more precise version of this result appears as Part (i) of Corollary 2.1 below). To use a result such as (1.13) to construct confidence bands for $\xi_p(x)$, we need to obtain the distribution of $\sup_{x\in K}|V_p(x)|$. This is in general quite impossible, as is explained in Section 2.

One strategy is to give the process V_p a representation that makes possible a simple way to simulate the distribution of $\sup_{x\in K}|V_p(x)|$ on a computer. (In fact, the covariance structure of V_p depends on unknown parameters. The representation makes it possible to simulate the distribution of $\sup_{x\in K}|\hat{V}_p(x)|$, where the covariance structure of $\hat{V}_p$ is a consistent estimate of the covariance structure of V_p.) We simulate many copies of $\sup_{x\in K}|\hat{V}_p(x)|$ and use them to obtain the required critical constants.

The construction of the bootstrap bands follows Hjort (1985a) and is straightforward. Below we give a sketch of it in the case of no censoring; a detailed description is given in Section 2.3. The Cox model (1.9) has two unknown parameters, S and β_0. We estimate those by $\hat{S}$ and $\hat{\beta}$, where $\hat{S}$ is defined by $\hat{S}(t) = \prod_{s\leq t}(1 - \hat{\Lambda}(ds))$. For individual i, we generate an artificial lifelength from the estimated model (1.9), i.e. generate a lifelength from the survival function $(\hat{S}(t))^{\exp(\hat{\beta}'x_i)}$. From these artificial lifelengths we calculate the function $\hat{\xi}_p^*(x), x \in K$ and form $\sqrt{n}(\hat{\xi}_p^*(x) - \hat{\xi}_p(x)), x \in K$. The bootstrap principle is that if $(\hat{S}, \hat{\beta})$ is close to (S, β_0), then the distribution of $\sup_{x\in K}\sqrt{n}|\hat{\xi}_p^*(x) - \hat{\xi}_p(x)|$ is close to that of $\sup_{x\in K}\sqrt{n}|\hat{\xi}_p(x) - \xi_p(x)|$. Thus, the required critical constants are obtained from the large number of copies of $\sup_{x\in K}\sqrt{n}|\hat{\xi}_p^*(x) - \hat{\xi}_p(x)|$.

Actually, we discuss several kinds of simulated-process bands and compare these and the bootstrap bands in terms of width and coverage probability in Monte Carlo studies. We provide theorems which state that both the simulated-process type bands and the bootstrap bands have asymptotically correct coverage probabilities.

This paper is organized as follows. Section 2 gives statements of the main theoretical results of the paper, and describes in detail our approach to constructing the simulated-process and the bootstrap bands, and "equal-precision" versions thereof analogous to the equal-precision bands developed by Nair (1984) and Hjort (1985b) in different contexts. Section 3 gives technical details on how the bands were formed and also details on how the computations were carried out. Section 4 reports Monte Carlo studies that compare all the bands. Perhaps surprisingly, these studies indicate that the simulated-process bands perform very well both in terms of width of the bands and in terms of coverage probability, even for moderate sample sizes. Section 4

also reports Monte Carlo studies that compare the three estimators $\bar{\xi}_p(x)$, $\tilde{\xi}_p(x)$ and $\hat{\xi}_p(x)$. Section 5 illustrates the various bands on the Stanford Heart Transplant Data. The Appendix provides proofs of the results stated in Section 2.

There are numerous examples of situations where one can establish weak convergence to a process whose distribution is intractable, and also depends on unknown parameters. In many cases one can, by a suitable transformation, reformulate the weak convergence result in such a way that the distribution of the limiting process is more tractable, and does not depend on unknown parameters. For example Efron's (1967) and Hall and Wellner's (1980) transformation of the Kaplan-Meier curve yield weak convergence to a Brownian Motion and a Brownian Bridge, respectively, and Nair's (1984) and Hjort's (1985b) rescaling transformations of the Kaplan-Meier curve and the Nelson-Aalen estimator, respectively, result in a weak convergence to a time transformed Ornstein-Uhlenbeck process. In more complicated situations (e.g. the Kaplan-Meier quantile process or the process $\hat{S}(\cdot)$ in the Cox model) such transformations are not available. We hope that the technique of using the simulated process will prove useful in these situations.

2 Notation and Theoretical Results

2.1 Notation and Basic Results

We consider a study involving n individuals where for individual i, X_i is a q-dimensional vector of covariates, Y_i is a lifetime, and C_i is a censoring time. We observe (T_i, δ_i, X_i), where $T_i = \min(Y_i, C_i)$ and $\delta_i = I(Y_i \leq C_i)$, $I(A)$ being the indicator function of the set A. Define the counting processes

$$N_i(t) = I(T_i \leq t; \delta_i = 1) \qquad t \geq 0 \tag{2.1}$$

and the processes

$$J_i(t) = I(T_i \geq t) \qquad t \geq 0. \tag{2.2}$$

In this notation, conditional on $X_i = x_i$, $i = 1, \ldots, n$, the partial likelihood of β_0 at time τ is

$$L(\beta, \tau) = \prod_{u \in [0,\tau]} \prod_{i=1}^{n} \left(\frac{J_i(u) \exp(\beta' x_i)}{\sum_{j=1}^{n} J_j(u) \exp(\beta' x_j)} \right)^{dN_i(u)}. \tag{2.3}$$

The maximum partial likelihood estimator of β_0 at time τ is the value $\hat{\beta} = \hat{\beta}(\tau)$ of β that maximizes $L(\beta, \tau)$. In practice, of course, one uses the value of β that maximizes the partial likelihood at time ∞; see the discussion in Section 4 of Andersen and Gill (1982) (henceforth AG). The "Nelson-Aalen" estimator of Λ is

$$\hat{\Lambda}(t) = \int_0^t \Big(\sum_{j=1}^{n} J_j(s) \exp(\hat{\beta}' x_j) \Big)^{-1} d\Big(\sum_{i=1}^{n} N_i(s) \Big). \tag{2.4}$$

The estimator $\hat{\Lambda}(t)$ increases only by jumps, which occur at the uncensored deaths. If we linearly interpolate $\hat{\Lambda}(t)$ we obtain $\hat{\Lambda}^c(t)$, Breslow's (1972,1974) estimator of Λ.

We remark that if β_0 is known to be 0, i.e. it is known that there is no covariate effect, and if there is no censoring, then the distribution corresponding to $\hat{\Lambda}(t)$ is the usual empirical distribution function. This is not the case for $\hat{\Lambda}^c(t)$.

The notation below follows closely that of AG, on whose results we rely heavily. For a q-vector $w = (w_1, \ldots, w_q)$, $w^{\bigotimes 2}$ denotes the $q \times q$ matrix whose $(i,j)^{th}$ entry is $w_i w_j$. Define

$$\begin{aligned}
S^{(0)}(\beta,t) &= \frac{1}{n}\sum_{l=1}^{n} J_l(t)\exp(\beta' x_l),\\
S^{(1)}(\beta,t) &= \frac{1}{n}\sum_{l=1}^{n} x_l J_l(t)\exp(\beta' x_l),\\
S^{(2)}(\beta,t) &= \frac{1}{n}\sum_{l=1}^{n} x_l^{\bigotimes 2} J_l(t)\exp(\beta' x_l), \qquad (2.5)\\
E(\beta,t) &= \frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)},
\end{aligned}$$

and

$$V(\beta,t) = \frac{S^{(2)}(\beta,t)}{S^{(0)}(\beta,t)} - (E(\beta,t))^{\bigotimes 2}.$$

We assume that the vector $(N_1, \ldots, N_n)$ forms a counting process with respect to some filtration $(\mathcal{F}_t,\ t \geq 0)$. (This is the case for example if the Y_i's are independent, and at time t it can be determined whether or not censoring has taken place, i.e. censoring is predictable; this does not require the C_i's to be independent). A good heuristic treatment of the concepts involved here is Gill (1984).

Fix $p^{(1)} \in (0,1)$ and let K be a bounded rectangle in $\mathcal{R}^q$. Define

$$\tau = \sup\Big\{\xi_p(x); p \in [0, p^{(1)}], x \in K\Big\}. \qquad (2.6)$$

Assume that for some $\epsilon > 0$, λ is continuous and positive on $[0, \tau + \epsilon]$ and assume Conditions A-D of AG(1982) with the interval $[0, \tau+\epsilon]$ in place of $[0,1]$. Condition B states that there exist scalar, vector, and matrix functions $s^{(0)}, s^{(1)}$, and $s^{(2)}$, respectively, of β and t, to which $S^{(0)}$, $S^{(1)}$, and $S^{(2)}$ converge uniformly over a neighborhood of β_0 and over $[0, \tau+\epsilon]$. Let

$$e(\beta,t) = \frac{s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)} \quad \text{and} \quad v(\beta,t) = \frac{s^{(2)}(\beta,t)}{s^{(0)}(\beta,t)} - (e(\beta,t))^{\bigotimes 2}. \qquad (2.7)$$

Then e is a q-vector and v is a matrix. Define $\Sigma = \Sigma(\tau)$ by

$$\Sigma = \int_0^{\tau} v(\beta_0,t) s^{(0)}(\beta_0,t)\, d\Lambda(t) \qquad (2.8)$$

and assume that Σ is positive definite. Also define

$$a(t) = \int_0^t (s^{(0)}(\beta_0,u))^{-1} d\Lambda(u) \quad \text{and} \quad b(t) = \int_0^t e(\beta_0,u)\, d\Lambda(u). \qquad (2.9)$$

Note that $b(t)$ is a q-vector. For R a bounded closed rectangle in $\mathcal{R}^m$, $C_m(R)$ will denote the space of real valued continuous functions defined on R with the sup norm topology.

Theorems 1, 2, and 3 below concern, essentially, the limiting distributions of the processes $\sqrt{n}(\bar{\xi}_p(x) - \xi_p(x))$, $\sqrt{n}(\tilde{\xi}_p(x) - \xi_p(x))$, and $\sqrt{n}(\hat{\xi}_p(x) - \xi_p(x))$ and of some of their relatives as (p, x) ranges over $[0, p^{(1)}] \times K$. These involve the space $C_{q+1} = C_{q+1}([0, p^{(1)}] \times K)$. Corollary 2.1 concerns the limiting distribution of these processes when p is fixed and x is scalar and involves the familiar space $C = C(K)$. For technical reasons, we find it necessary to work with $C_m(R)$ instead of the "Skorohod space" $D_m(R)$ (see the Appendix).

Let $\left(\hat{S}(\cdot|x)\right)^l$ be the function obtained by linearly interpolating $\hat{S}(\cdot|x)$, and let $\hat{\xi}_p^c(x)$ be obtained by solving for t in the equation $\left(\hat{S}(t|x)\right)^l = 1 - p$. This $\hat{\xi}_p^c(x)$ is continuous. Details on the exact definition of $\hat{\xi}_p^c(x)$ appear in Section 4. Here, we simply note that all the estimators are asymptotically equivalent uniformly in $p \in [0, p^{(1)}], x \in K$; this is made clear in the proof of Theorem 1. We shall use $\hat{\xi}_p(x)$ to refer to either $\hat{\xi}_p(x)$ or $\hat{\xi}_p^c(x)$ except in the exact statements of the theorems and in the proofs in the Appendix.

In the results below, $W(\cdot)$ denotes a standard Brownian Motion on $[0, \infty)$ and Z a q-dimensional random vector that is normally distributed with mean 0 and identity as covariance matrix, and is independent of $W(\cdot)$. We use the notation $\pi = \log(1-p)$; also, $\xrightarrow{d}$ denotes convergence in distribution, and $\xrightarrow{pr.}$ denotes convergence in probability.

Theorem 1 *Under the assumptions stated above, as $n \to \infty$*

$$\sqrt{n}(\hat{\xi}_p^c(x) - \xi_p(x)) \xrightarrow{d} V(p, x)$$

in $C_{q+1}\left([0, p^{(1)}] \times K\right)$, where

$$V(p,x) = \frac{W\left(a(\xi_p(x))\right)}{\lambda(\xi_p(x))} + \left(\frac{\left(b(\xi_p(x)) + \pi x \exp(-\beta_0' x)\right)' \Sigma^{-1/2}}{\lambda(\xi_p(x))}\right) Z.$$

Theorem 1 indicates that the variance of $V(p, x)$ is

$$\sigma^2(p,x) = \frac{a(\xi_p(x))}{(\lambda(\xi_p(x)))^2} + \left(\frac{b(\xi_p(x)) + \pi x \exp(-\beta_0' x)}{\lambda(\xi_p(x))}\right)' \Sigma^{-1} \left(\frac{b(\xi_p(x)) + \pi x \exp(-\beta_0' x)}{\lambda(\xi_p(x))}\right) \tag{2.10}$$

From (2.10) we see that to estimate $\sigma(p, x)$ we need to estimate the functions $a(\cdot), b(\cdot)$ and the matrix Σ; in addition, we shall need to estimate $\lambda(\cdot)$. For $a(\cdot), b(\cdot)$ and Σ, we introduce the natural estimates

$$\begin{aligned} \hat{a}(t) &= \int_0^t (S^{(0)}(\hat{\beta}, u))^{-1} d\hat{\Lambda}(u), \\ \hat{b}(t) &= \int_0^t E(\hat{\beta}, u) d\hat{\Lambda}(u), \end{aligned} \tag{2.11}$$

and

$$\hat{\Sigma} = \int_0^{\tau} V(\hat{\beta}, t) S^{(0)}(\hat{\beta}, t) d\hat{\Lambda}(t).$$

The estimation of λ is somewhat more difficult. Possibilities include methods based on splines (Whittemore and Keller, 1986) and those based on kernel smoothers. Kernel smoothers are computationally convenient, and in addition their asymptotic properties in the present context have already been studied by Ramlau-Hansen (1983). To describe them, let R be a function of bounded variation with support on $[-1, 1]$, and whose integral is 1, and let $\{b_n\}$ be a sequence of positive constants such that as $n \to \infty$, we have $b_n \to 0$ and $nb_n^2 \to \infty$. Define the kernel estimate of $\lambda(\cdot)$ by

$$\hat{\lambda}(t) = \frac{1}{b_n} \int_0^{\infty} R\Big(\frac{t-s}{b_n}\Big) d\hat{\Lambda}(s). \tag{2.12}$$

The specific choices of R and $\{b_n\}$ are discussed, in the context of ordinary density estimation, in Silverman (1986, pp. 40–72). See our discussion in Section 4. Having specified a choice of R and $\{b_n\}$ we may define an estimate $\hat{\sigma}(p, x)$ of $\sigma(p, x)$ by substituting in (2.10) $\hat{a}(\cdot), \hat{b}(\cdot), \hat{\Sigma}, \hat{\xi}_p(x), \hat{\beta}$, and $\hat{\lambda}(\cdot)$ for $a(\cdot), b(\cdot), \Sigma, \xi_p(x), \beta$, and $\lambda(\cdot)$, respectively. The following lemma gives uniform consistency of this estimator.

Lemma 2.1 *Under the conditions stated above, as $n \to \infty$*

(*i*) $\sup_{0 \le t \le \tau+\epsilon} |\hat{a}(t) - a(t)| \xrightarrow{pr.} 0$

(*ii*) $\sup_{0 \le t \le \tau+\epsilon} |\hat{b}(t) - b(t)| \xrightarrow{pr.} 0$

(*iii*) $\hat{\Sigma} \xrightarrow{pr.} \Sigma$

(*iv*) $\sup_{p \in [0, p^{(1)}], x \in K} |\hat{\xi}_p(x) - \xi_p(x)| \xrightarrow{pr.} 0$

(*v*) *for any* $\eta > 0$,
$\sup_{\eta \le t \le \tau} |\hat{\lambda}(t) - \lambda(t)| \xrightarrow{pr.} 0$

(*vi*) *for any* $p^{(0)} > 0$,
$\sup_{p \in [p^{(0)}, p^{(1)}], x \in K} |\hat{\sigma}(p, x) - \sigma(p, x)| \xrightarrow{pr.} 0$

(*vii*) *Let $\hat{a}^c$ and $\hat{b}^c$ be given by (2.11), except that $\hat{\Lambda}$ is replaced by the version of $\hat{\Lambda}$ obtained by linearly interpolating $\hat{\Lambda}$. Then the conclusion of Part (vi) is still true for the version $\hat{\sigma}^c(p, x)$ of $\hat{\sigma}(p, x)$ obtained by using $\hat{a}^c, \hat{b}^c$ and $\hat{\xi}_p^c(x)$ instead of $\hat{a}, \hat{b}$ and $\hat{\xi}_p(x)$.*

Suppose that we were able to obtain c_α, the $(1-\alpha)^{th}$ quantile of the distribution of $\sup_{p \in [p^{(0)}, p^{(1)}], x \in K} |V(p, x)|$. We would then have

$$\lim_{n \to \infty} P\left\{\hat{\xi}_p(x) - \frac{c_\alpha}{\sqrt{n}} \le \xi_p(x) \le \hat{\xi}_p(x) + \frac{c_\alpha}{\sqrt{n}} \quad \text{for all} \quad p \in [p^{(0)}, p^{(1)}], x \in K\right\} \to 1 - \alpha.$$

i.e. a $(1-\alpha) \times 100\%$ asymptotic confidence band for $\xi_p(x)$ is $\hat{\xi}_p(x) \pm c_\alpha/\sqrt{n}$. This Kolmogorov-Smirnov type of confidence band suffers the defect that it has constant

width. One would want the band to be narrower at those values of (p, x) where the variance of $\hat{\xi}_p(x)$ is small. In the context of forming confidence bands for a survival function in the random censorship model of survival analysis, Nair (1984) proposed a confidence band with the property that the width of the band at a given point is proportional to the estimated standard deviation at that point. He called the band "equal-precision band". A similar idea was proposed by Hjort (1985b) for estimation of cumulative hazard rates. We proceed along a similar route.

Theorem 2 *Under the regularity conditions stated above, as $n \to \infty$*

$$\sqrt{n}\Big(\frac{\hat{\xi}_p^c(x) - \xi_p(x)}{\hat{\sigma}^c(p, x)}\Big) \xrightarrow{d} \frac{V(p, x)}{\sigma(p, x)}$$

in $C_{q+1}\Big([p^{(0)}, p^{(1)}] \times K\Big)$.

2.2 The Simulated Process Bands

Our procedure for using Theorem 2 to construct confidence bands amounts to estimating the parameters of the process

$$L(p, x) = \frac{V(p, x)}{\sigma(p, x)}$$

and then generating a process $\hat{L}(p, x)$ with those estimated parameters. To describe it in more detail, note that we may write

$$L(p, x) = c(p, x)W\Big(a(\xi_p(x))\Big) + d(p, x)Z \tag{2.13}$$

where

$$\begin{aligned} c(p, x) &= \Big[a(\xi_p(x)) + \Big(b(\xi_p(x)) + \pi x \exp(-\beta_0' x)\Big)'\Sigma^{-1}\Big(b(\xi_p(x)) + \pi x \exp(-\beta_0' x)\Big)\Big]^{-1/2} \\ &\text{and} \\ d(p, x) &= \Big(b(\xi_p(x)) + \pi x \exp(-\beta_0' x)\Big)'\Sigma^{-1/2} \\ &\quad \Big[a(\xi_p(x)) + \Big(b(\xi_p(x)) + \pi x \exp(-\beta_0' x)\Big)'\Sigma^{-1}\Big(b(\xi_p(x)) + \pi x \exp(-\beta_0' x)\Big)\Big]^{-1/2} \end{aligned} \tag{2.14}$$

Let $\hat{c}(p, x)$ and $\hat{d}(p, x)$ be the estimates of $c(p, x)$ and $d(p, x)$ obtained by replacing all the unknowns by their continuous estimates. We generate the process

$$\hat{L}(p, x) = \hat{c}(p, x)W\Big(\hat{a}^c(\hat{\xi}_p^c(x))\Big) + \hat{d}(p, x)Z \tag{2.15}$$

where $W(\cdot)$ is a standard Brownian motion, Z is an independent q-variate standard normal variable and $(W(\cdot), Z)$ is independent of all parameter estimates, and calculate $M = \sup_{p\in[p^{(0)}, p^{(1)}], x\in K} |\hat{L}(p, x)|$. To estimate $s_\alpha^{(n)}$, the $(1-\alpha)^{th}$ quantile of the distribution of M, we repeat the above step independently n_0 times, for some

large number n_0, obtaining iid copies $M_1, M_2, \ldots, M_{n_0}$ of M, and take the empirical $(1-\alpha)^{th}$ quantile of $M_1, \ldots, M_{n_0}$ as an approximation to $s_\alpha^{(n)}$. This produces the band

$$\hat{\xi}_p(x) \pm s_\alpha^{(n)} \hat{\sigma}(p,x)/\sqrt{n}. \tag{2.16}$$

Theorem 3 *Under the regularity conditions stated above, as $n \to \infty$*

$$P\left\{\sqrt{n}\left|\frac{(\hat{\xi}_p(x) - \xi_p(x))}{\hat{\sigma}(p,x)}\right| \leq s_\alpha^{(n)} \text{ for all } p \in [p^{(0)}, p^{(1)}], x \in K\right\} \to 1-\alpha$$

Thus the band (2.16) has asymptotic coverage probability $1-\alpha$.

2.3 The Bootstrap Bands

Suppose that the censoring variables can be thought of as being iid from some survival function R_C, and also independent of the survival times. If we view the X_i's as fixed at x_i, the Cox model is then specified by the triple (S, β_0, R_C). Let $\hat{R}_C$ be the Kaplan-Meier estimate of R_C. A natural way to resample (see Hjort (1985a)) is to generate artificial data as follows.

For each $i = 1, \ldots, n$

1 Generate $Y_i^* \sim (\hat{S}(\cdot))^{\exp(\hat{\beta}' x_i)}$

2 Generate $C_i^* \sim \hat{R}_C$ (all observations independent)

3 Form $T_i^* = \min(Y_i^*, C_i^*)$ and $\delta_i^* = I(Y_i^* \leq C_i^*)$

This gives one artificial data set (T_i^*, δ_i^*, x_i), $i = 1, \ldots, n$ from which we calculate $\hat{\Lambda}^*$ and $\hat{\beta}^*$, and we use those to construct the function $\hat{\xi}_p^*(x)$ (and also $\hat{\xi}_p^{*c}(x)$, for the technical statement of Theorem 4). This is repeated independently a large number of times.

Efron and Tibshirani (1986) also discuss the scheme of bootstrapping by resampling from the triples (T_i, δ_i, X_i), $i = 1, \ldots, n$.

Suppose that we wish to estimate the variability of some estimate, in our case $\hat{\xi}_p(x)$. The first method mentioned is appropriate for estimating the conditional variance of $\hat{\xi}_p(x)$ given the X's. The second method is appropriate for estimating the unconditional variance of $\hat{\xi}_p(x)$, i.e. averaging over the marginal distribution of the covariates and of the censoring variables. If the distribution of the X_i's does not depend on the unknown parameters S and β_0 then the usual ancillarity arguments point to the conditional variance as the "right" quantity to estimate. This is our point of view, and the results below pertain to the first method of resampling.

This situation is closely connected to bootstrapping in linear regression models, where one can bootstrap by resampling from the pairs (responses, covariates), or one can bootstrap by resampling from the residuals: see Freedman (1981). Many of the comments in the discussion paper Wu (1986) are relevant here.

To show that the bands are asymptotically valid we need to show that for large n, the distribution of $\sqrt{n}(\hat{\xi}_p^*(x) - \hat{\xi}_p(x))$ is close to that of $\sqrt{n}(\hat{\xi}_p(x) - \xi_p(x))$. The technical statement of this fact involves the notion of "weak convergence in probability", and is given as Theorem A.1 in the Appendix.

To construct the bands based on the bootstrap, let $w^* = \sup_{p\in[0,p^{(1)}],x\in K} \sqrt{n}|\hat{\xi}_p^*(x) - \hat{\xi}_p(x)|/\hat{\sigma}(p,x)$. Obtain a large number of copies of w^*, say $w_1^*, \ldots, w_B^*$ and let b_α^n be the $(1-\alpha)^{th}$ quantile of the empirical distribution of $w_1^*, \ldots, w_B^*$.

Theorem 4 *Assume the conditions of Theorem 1. If sampling is carried out via steps 1,2, and 3, then the band*

$$\hat{\xi}_p(x) \pm b_\alpha^{(n)} \hat{\sigma}(p,x)/\sqrt{n}$$

has asymptotic coverage probability $1-\alpha$.

Remark: In our simulation studies, we have used the bootstrap estimate of standard error instead of $\hat{\sigma}(p,x)$. That is, before bootstrapping to get the critical constants as described above, the standard error function of the process is estimated using a separate set of bootstrap samples. This does not require the estimation of the hazard rate, but is much more computationally intensive.

2.4 Extensions and Special Cases

There are several corollaries and extensions to Theorems 1–4 that could be stated. Examples include the following.

a The case of p fixed at p_0, and $x_1, \ldots, x_l$ fixed while $x_{l+1}, \ldots, x_q$ vary freely. Weak convergence then takes place in the space C_{q-l}.

b The case of p fixed at p_0, and x varying freely through a smooth subset S of the rectangle K. An example of this arises if the model is a polynomial regression in the scalar u, i.e.

$$\lambda(t|u) = \lambda(t)\exp(\beta_1 u + \beta_2 u^2 + \ldots + \beta_d u^d),$$

and then S is simply a line through K. Weak convergence takes place in the space of continuous functions defined on S. An example of this is the Stanford Heart Transplant Data (see Section 5) in which we used a second degree polynomial.

For the sake of reference we state (Corollary 2.1 below) our results for the simple but important case of p fixed at p_0 and of scalar x varying over the set $K = [K_1, K_2]$. The usual assumptions are in force (in particular we assume that λ is continuous and positive on $[0, \sup_{K_1\le x\le K_2} \xi_{p_0}(x) + \epsilon]$). We denote $V(p_0,x)$ and $\sigma(p_0,x)$ by $V_{p_0}(x)$ and $\sigma_{p_0}(x)$, respectively; $s_\alpha^{(n)}(p_0)$ is the $(1-\alpha)^{th}$ quantile of the distribution of $\sup_{K_1\le x\le K_2} |\hat{L}(p_0,x)|$, and $b_\alpha^{(n)}(p_0)$ is the corresponding quantity for the bootstrap process.

Corollary 2.1 *Fix p at p_0 and let the scalar x vary over $[K_1, K_2]$. Then as $n \to \infty$*

(i) $\sqrt{n}(\hat{\xi}^c_{p_0}(x) - \xi_{p_0}(x))/\hat{\sigma}_{p_0}(x) \xrightarrow{d} V_{p_0}(x)/\sigma_{p_0}(x)$ *in* $C[K_1, K_2]$

(ii) *The bands*

$$\hat{\xi}_{p_0}(x) \pm s_\alpha^{(n)}(p_0)\hat{\sigma}(p_0, x)/\sqrt{n}$$

and

$$\hat{\xi}_{p_0}(x) \pm b_\alpha^{(n)}(p_0)\hat{\sigma}(p_0, x)/\sqrt{n}$$

each have asymptotic coverage probability $1 - \alpha$.

3 On Computation of the Bands

In this section we begin by discussing two topics pertaining to the simulated-process bands—estimating the hazard rate, and forming another type of simulated-process band by using the asymptotic distribution of $\log \hat{\xi}_p(x)$. Then we remark on papers of Hall on how many bootstrap replications are needed in forming confidence intervals. Finally we give information on the computers and programs which were used in carrying out the studies.

The simulated-process equal-precision band is given in (2.16), (2.13), and (2.14), with formulas for estimates of the parameters of the limiting process given by (2.11) and (2.12). The kernel estimate of the hazard rate, however, is not fully defined by (2.12); the kernel function and the bin width must be specified. It is well known that, whereas the choice of kernel function is not critical, the choice of bin width is. Silverman (1986) emphasizes in particular that the fixed-width kernel estimator is defective when applied to long-tailed distributions: "Because the window width is fixed across the entire sample, there is a tendency for spurious noise to appear in the tails of the estimates; if the estimates are smoothed sufficiently to deal with this, then essential detail in the main part of the distribution is masked" (Silverman, 1986, p. 18). In Chapter 5 of his book, he discusses the adaptive kernel method, which "is based on the common-sense notion that a natural way to deal with long-tailed densities is to use a broader kernel in regions of low density" (ibid., p. 100). Silverman deals only with density estimates based on iid samples, and in our problem we need an estimate of the hazard rate from non-iid data; however, his remarks are pertinent here since in survival analysis, densities very often *are* long-tailed, and in addition, as time goes on, there is less accurate information on the hazard rate available from the data due to censoring as well as earlier deaths. In a paper primarily on asymptotics for the kernel estimator of the hazard rate in the random censorship model of survival analysis, Ramlau-Hansen (1983) illustrates the kernel method on a data set for which he chooses in an ad hoc manner three intervals of the time axis and different bin widths for each interval, such that the bin width increases with time. In this paper we use the biweight kernel

$$K(t) = \frac{15}{16}(1 - t^2)^2 \qquad |t| < 1 \tag{3.1}$$

We have followed Ramlau-Hansen in allowing bin widths to increase as time increases, but we need an automatic choice of bin widths for the simulation studies. Noting Ramlau-Hansen's criteria for uniform consistency of the kernel estimator of the hazard rate, that $b_n \to 0$ and $nb_n^2 \to \infty$ as $n \to \infty$, the bin widths were taken to be inversely proportional to $n^{1/3}$. Then with the data from one simulation, a subjective choice of time intervals and bin widths was made. This choice was used for all the simulations. This particular procedure by no means leads to an optimal choice of bin width, but optimality is not our goal here—we want to see how well the simulated-process band performs with a merely *reasonable* choice of bin width. Another method to obtain an automatic choice of bin width would be to modify the algorithm given by Tanner (1984), which is for estimation of the hazard rate function in the random censorship model with no covariates. This algorithm applies cross-validation to the log-likelihood function to select three smoothing parameters.

Preliminary studies indicated that a log transformation greatly improved the coverage probability of the simulated-process band in many cases. For the problem of forming a confidence band for the cumulative hazard function $\Lambda(t)$ in the random censorship model without covariates, Bie, Borgan, and Liestol (1984) found that bands based on transformations of the Nelson-Aalen estimator $\hat{\Lambda}(t)$ had much closer to nominal coverage probability than bands based on the untransformed $\hat{\Lambda}(t)$, in several Monte Carlo studies with exponential and Weibull survival times and exponential and uniform censoring times. Kalbfleisch and Prentice (1980, p. 15) mentioned that parameter transformations do not seem to have been considered for survival analysis inference problems, although parameter transformations can improve the adequacy of normal approximations and avoid the problem of impossible values occurring in a confidence interval or band.

In the present situation, an argument for the log transformation is that it is a variance-stabilizing transformation in a special case. When the variance of a univariate observation X is some function of its mean, a variance-stabilizing transformation is a function g such that the variance of $g(X)$ does not depend on the mean. The log transformation is variance-stabilizing when the standard deviation is proportional to the mean. In the present situation, consider a single, fixed value of the covariate x. Call the function g variance-stabilizing at x if the limiting variance of $g(\hat{\xi}_p(x))$ does not depend on $\xi_p(x)$. For fixed x, this limiting variance depends on the covariate, lifetime, and censoring distributions, as well as on the Cox model regression parameter β_0. To simplify matters, consider the following particular situation: a single covariate distributed uniformly on (0,1), and exponential distributions for both the lifetime and censoring variables. A formula for the limiting variance of $\hat{\xi}_p(x)$ can be written down. This formula did not turn out to have a nice form: it involves integrals which must be evaluated numerically. Following through to the end of this argument, the calculations were done for several specific choices of the parameters of the exponential distributions of the lifetime and censoring variables, in order to plot the limiting variance of $\hat{\xi}_p(x)$ as a function of $\xi_p(x)$. In each case, the resulting curve appeared very close to quadratic, suggesting the choice of the log transformation.

This argument is by no means decisive: What we actually want is a normalizing transformation, which is not necessarily the same as a variance-stabilizing transformation, but it is not known if such a transformation exists for this problem. (Efron (1982) discusses the problem of finding normalizing transformations in the case of a one-parameter family of distributions.) Our argument merely serves to *suggest* that we investigate the performance of the log-transformed bands through Monte Carlo studies.

A standard delta-method argument is used, via Theorem 2 in Section 2, to form the log-transformed equal-precision band. Let g be a differentiable, monotone function on $[\xi_p(x_1), \xi_p(x_2)]$. Then $\sqrt{n}\big(g(\hat{\xi}_p(x)) - g(\xi_p(x))\big) \xrightarrow{d} g'(\xi_p(x))V(p,x)$ on $C[\xi_p(x_1), \xi_p(x_2)]$, and so the standard deviation function of $\sqrt{n}g(\hat{\xi}_p(x))$ is $g'(\xi_p(x))\sigma(p,x)$. To get the equal-precision band for $g(\xi_p(x))$, write

$$\frac{\sqrt{n}\big(g(\hat{\xi}_p(x)) - g(\xi_p(x))\big)}{g'(\hat{\xi}_p(x))\hat{\sigma}(p,x)} \xrightarrow{d} \frac{V(p,x)}{\sigma(p,x)}$$

on $C[\xi_p(x_1), \xi_p(x_2)]$.

So the $100(1-\alpha)\%$ confidence band for $g(\xi_p(x))$ is $g(\hat{\xi}_p(x)) \pm g'(\hat{\xi}_p(x))\hat{\sigma}(p,x)s_\alpha/\sqrt{n}$, where s_α is the critical constant for the equal-precision band. For a given monotone function, this band may be converted into a $100(1-\alpha)\%$ band for $\xi_p(x)$. For the log transformation the band is

$$\hat{\xi}_p(x)\exp\Big\{\pm\frac{1}{\hat{\xi}_p(x)}\frac{\hat{\sigma}(p,x)}{\sqrt{n}}s_\alpha\Big\}.$$

The bootstrap resampling scheme and equal-precision band are described in Section 2.3. In determining number of bootstrap replications, we noted two papers by Hall (1986a,b). In the situation of iid vector observations in which the parameter of interest is a function of the mean vector, he gives expansions for the probability of coverage of one-sided bootstrap percentile-t intervals. He then uses these results to show that in the situation he considers, a very small number of bootstrap replications is adequate to get very close to the true bootstrap coverage probability, that is, the coverage probability with an infinite number of bootstrap replications. He actually shows that for any fixed finite number of bootstrap samples, the worst departure of actual coverage probability from nominal coverage probability, over coverage probabilities for $\alpha \in (0,1]$, is less than or equal to the worst departure for an infinite number of bootstrap replications. He cautions that his result does not mean that one should use a small number of replications in practice, for such bootstrap intervals will be wider than those derived from a large number of bootstrap replications. Hall's theory is quite specialized: His situation does not include estimation of the parameters in the Cox model, and he does not consider confidence bands. Monte Carlo studies indicated, however, that the coverage probability of our bootstrap bands when only 19 bootstrap replications were used was usually surprisingly close to the coverage probability when 1000 bootstrap replications were used. Except for these papers, it

would not have occurred to us that coverage probability could be so little affected by taking a very small number of bootstrap replications. Bootstrapping in the Cox model requires a large amount of CPU time, and using a small number of bootstrap replications in the debugging stage was an easy way to save time. We mention this in the hope that using only a small number of bootstrap samples will prove useful in the early stages of other empirical studies of the bootstrap, even when Hall's conditions are not met.

No transformation of the bootstrap bands is considered here. Since the simulated-process bands require fairly substantial computational effort, a natural question was, "Why not just bootstrap?" A chief virtue of the bootstrap is that it is automatic, that is, it requires no asymptotic theory to derive, nor any special tricks to make it work, and here we just want to see how well this automatic, or untransformed, bootstrap does compared with the simulated-process method.

Computations were carried out using FORTRAN programs compiled with the f77 Unix compiler on the Florida State University Statistics Department network of Sun computers. The computation of the bootstrap band on a sample of size 80 required almost 2 minutes of CPU time on a Sparcstation 1; computation of the simulated-process band required 7 seconds. A study of bootstrap bands for $n=80$ with 5,000 simulations, split between two Sparcstations, required about 3.5 days to complete. Programs and subroutines for Cox model fitting were adapted for simulation purposes from the programs given in Kalbfleisch and Prentice (1980, Appendix 3). The algorithm uses the usual Peto approximation for ties. For the simulation studies, pseudo-random uniformly distributed variables were obtained with a FORTRAN implementation of the "universal random number generator" of Marsaglia and Zaman (1987). This random number generator combines a lagged-Fibonacci sequence with a simple arithmetic sequence. It has a period of about 2^{144} and satisfies stringent tests of randomness. It was designed so that in all CPU's with at least 16-bit integer arithmetic, given the same four starting seeds, the same sequence of uniform variables is produced. Normal and exponential variables are generated by the ziggurat method of Marsaglia and Tsang (1984).

Fortran programs to calculate all the bands are available from the authors on request.

4 Simulation Studies

Here we report results of a Monte Carlo study comparing the three methods of forming bands, and of Monte Carlo studies comparing the three estimators of $\xi_p(x)$ which were introduced in Section 1.

Table 1 gives a summary of results of a Monte Carlo study of simulated-process bands with 10,000 simulations, and of another Monte Carlo study of bootstrap bands with 5,000 simulations. Results for a single simulation situation are given here. Results for other situations were similar to these. In the situation reported on here, there is one covariate x which is evenly spaced on the interval [0,1]. The lifetime dis-

tribution is Exponential(1), and the Cox regression parameter β_0 is 1. The censoring distribution is also exponential, with mean set to give the desired degree of censoring. In these studies, $n = 80$ and the mean of the censoring variable is 2.49, which gives 20% censoring. The number of bootstrap samples to estimate the standard deviation function of the process is 200; the number of bootstrap samples to obtain the critical constants is 599.

In this situation, the bootstrap bands have somewhat higher coverage probability but are much wider than the simulated-process bands. The log-transformed simulated-process bands have slightly lower coverage probability and are slightly wider than the standard simulated-process bands. As can be seen from the table, the simulated-process equal-precision bands did very well. The mean widths appear large for all the methods, compared to the mean survival time, which ranges from 1.0 to 2.7 for $x \in [0,1]$. However, these means include the widths at the left and right extremes of the covariate values, where the bands can be very wide, and therefore it is helpful to look at the mean critical constants, which for equal-precision bands, show how much precision is lost by going from confidence interval to confidence band. For the simulated-process band, the mean critical constants of 2.60 and 2.32 compare favorably to 1.96 and 1.65 for 95% and 90% confidence intervals.

To better understand the relative merits of the two methods, we felt that it would be useful to see how they do on the more basic problem of forming confidence intervals for $\xi_p(x)$. The simulated-process 95% confidence interval for $\xi_p(x)$ is just $\hat{\xi}_p(x) \pm c_{.025}\hat{\sigma}(p,x)/\sqrt{n}$ where $c_{.025}$ is the .975 quantile of the standard normal distribution. Two types of bootstrap confidence intervals for $\xi_p(x)$ were formed. The first kind was suggested by the bootstrap confidence band procedure, and we studied it partly as further investigation into performance of the bootstrap band procedure. First form many bootstrap samples, and get values $\hat{\xi}_p^*(x) - \hat{\xi}_p(x)$ from each bootstrap sample. Take the .025 and .975 quantiles of this bootstrap distribution of $\hat{\xi}_p^*(x) - \hat{\xi}_p(x)$, c_L and c_U. The 95% confidence interval for $\xi_p(x)$ is then $(\hat{\xi}_p(x) - c_U, \hat{\xi}_p(x) - c_L)$. (Remark: In forming the bootstrap confidence bands, the standard error function of the process, $\sigma(p,x)$, is estimated before carrying out the bootstrap procedure for getting critical constants, and so the standard error at a single x is constant throughout this second stage of bootstrapping. Thus consideration of the bootstrap distribution of $(\hat{\xi}_p^*(x) - \hat{\xi}_p(x))/\hat{\sigma}(p,x)$ yields the same intervals as just described.) The second kind of bootstrap interval was just the most basic type, the standard percentile interval. For eleven values of the covariate x, evenly spaced on the interval [0,1], the simulated-process intervals and bootstrap percentile intervals had coverage probabilities very close to 95%. However, the coverage probabilities of the first kind of bootstrap interval ranged from 85% to 87% over the 11 x-values. Computation of average critical constants c_L and c_U for this type of bootstrap interval showed that the bootstrap distribution $\hat{\xi}_p^*(x) - \hat{\xi}_p(x)$ was skewed right. Somehow, in spite of the low coverage probability of the bootstrap individual confidence intervals, the features of the skewed bootstrap distribution and the taking the sup over a range of x's combine to give conservative bootstrap bands.

The main conclusion of these studies is that the simulated-process equal-precision bands perform surprisingly well, especially considering the moderate sample size used here. It was not clear to us before the Monte Carlo studies that simulation of the estimated process would be a feasible approach to obtaining critical constants for confidence bands in general; in addition, the Cox model is quite complicated, and we didn't know how estimation of the hazard rate would affect the performance of the bands.

In other Monte Carlo studies not reported here, in which the bin widths for the kernel estimator of hazard rate were deliberately chosen to be not good, the simulated-process method often had low coverage probability. The log transformed bands were then a noticeable improvement, and of course, the bootstrap bands were not affected. They have been conservative in all our studies.

Table 2 gives the results of several Monte Carlo studies comparing the three estimators of median survival time. Continuous versions of the estimators are used, and for these studies the aim is to eliminate the choice of smoothing method as a factor in the comparative performance of the estimators. Breslow's piecewise linear smooth of the cumulative hazard function is used for the first estimator, $\bar{\xi}_p^c(x)$, and to correspond to it, piecewise exponential smoothing of the product integral estimator of the survival function is used for the second and third estimators, $\tilde{\xi}_p^c(x)$ and $\hat{\xi}_p^c(x)$. The product integral estimator of the survival function can be negative. When this occurred in these studies, the estimator of the median of the survival distribution was found by linear interpolation between the largest death time for which $\hat{S} > .5$ and 0.

The six factors determining the simulation situation are the survival distribution, type of censoring distribution, percent censoring, covariate distribution, sample size, and value of the Cox regression parameter β_0. In these studies, the censoring distribution is always exponential, and the parameter β_0 is 1. There are two levels of each of the other four factors. The two choices of survival distribution are Exponential(1) and Weibull with shape parameter 2.0 and scale parameter 1.35. The two choices of percent censoring are 10% and 20%. The x's are fixed in the studies. For one level of the covariate factor, the covariate is taken evenly spaced on the interval (0,1); for the other level, the interval is (-1,1). The sample sizes are 20 and 50. The table reports the mean integrated squared error. By "integrated squared error" is meant average squared error over the grid of x values considered. The number of simulations in these studies is 10,000.

The overall conclusions are that the two product integral estimators $\hat{\xi}_p^c(x)$ and $\tilde{\xi}_p^c(x)$ outperform the original estimator $\bar{\xi}_p^c(x)$ in terms of mean integrated squared error, and that there is not much difference between $\hat{\xi}_p^c(x)$ and $\tilde{\xi}_p^c(x)$, although on the whole, $\hat{\xi}_p^c(x)$ does somewhat better. In some rows of Table 2, $\hat{\xi}_p^c(x)$ has much smaller mean integrated squared error than either $\bar{\xi}_p^c(x)$ or $\tilde{\xi}_p^c(x)$, see especially rows 3, 7, and 15. In each of these simulations, the covariate x ranges from -1 to 1; there is less information on median survival time for the small covariate values, since they tend to produce large survival times. Also, these simulations had a small sample size of

20. In this situation, there is a nonnegligible probability that for some small values of x, the estimator will have to be defined by extrapolation beyond the last observed death time, with potential for large outliers. Therefore, for these three rows, the 5% trimmed means of the integrated squared errors are also reported in the table. The product integral estimator $\hat{\xi}_p^c(x)$ still shows substantial improvement over the other two estimators. This apparent improvement of $\hat{\xi}_p^c(x)$ over the other two estimators here may be due merely to the particular choice of smoothing algorithm which was used; therefore, we do *not* claim that it provides as large an improvement as the tabled numbers indicate. The empirical evidence from these studies gives support for the product integral method beyond the theoretical arguments for it.

5 Illustration of the Bands on Real Data

The Stanford Heart Transplant Data have appeared frequently in the statistics literature. For a thorough analysis and discussion, see the discussion paper of Aitkin, Laird, and Francis (1983). Here we do not attempt to add any insights on this data set, but only to illustrate the methods of forming bands on a well-known case. We use the 1980 version of the data, as given in Miller and Halpern (1982). We also follow the Miller and Halpern Cox model analysis of the data, in which the final model was a quadratic regression of $\log_{10}$ survival time on the covariate age for the 152 patients with complete records who had survived at least 10 days. Figure 1 shows the three confidence bands for the median log survival time. The kernel estimate of the hazard rate, needed by the two types of simulated-process bands, used constant bin width of .27, and is pictured in Figure 2. The bootstrap bands are almost always wider than the simulated-process bands over the entire range of the covariate age, except at the far left, where both types of bands are so wide that the clear message is that there is no information on survival time distribution for these very young ages. The bootstrap band is wider than the simulated-process bands in the right half of the graph, but it is more appealing than the other two bands here because it is smoother. Its wideness goes along with its performance in the Monte Carlo studies. The unevenness of the simulated-process bands in the right end of the graph corresponds to bumps in the hazard rate estimator. The log-transformed bands are noticeably different from the untransformed equal-precision bands at the left end of the range of the covariate; otherwise these two bands are similar. To give an idea of the width of the bands, upper and lower endpoints of the 95% bands and individual confidence intervals are given in Table 3 at two values of age, 38.5 and 48.7 years. Considering the simulated-process equal-precision results only, we see that although the individual intervals barely overlap, the bands at the two points overlap considerably, so that one can't infer from the band a definite difference in median survival time for the two ages. Finally, we note that since the band has probability .95 of containing the true median survival time simultaneously for all x, it allows one to "snoop" through all the x values looking for interesting significant differences.

Appendix: Proofs

Theorems 1, 2, 3, and A.1 involve the space $C_{q+1}(R)$. We would like to prove that weak convergence takes place in the space $D_{q+1}([0,p^{(1)}] \times K)$ of functions defined on $[0,p^{(1)}] \times K$ which are "continuous from above, with limits from below" with the "Skorohod topology"; see e.g. Neuhaus (1971). However, none of the processes $\bar{\xi}_{\cdot}(\cdot)$, $\tilde{\xi}_{\cdot}(\cdot)$, and $\hat{\xi}_{\cdot}(\cdot)$ need be in this space. To see this, let's look at the case of p fixed and of scalar x. Consider for example $\hat{\xi}_{\cdot}(\cdot)$. Recall that

$$\hat{\xi}_p(x) = \hat{S}^{-1}\Big((1-p)^{\exp(-\hat{\beta}'x)}\Big).$$

This is the composition of $\hat{S}^{-1}$ (which is right continuous with left limits by construction) with a decreasing function of x. The result is a left continuous function with right limits, and is not in $D(K)$.

So we are forced to redefine our processes to make them continuous; we may then just as well work with the simpler spaces C_{q+1} and C.

Proof of Theorem 1

Our proof is based on a general result concerning weak Bahadur representations for quantile processes (Proposition A.1 below) and on AG's weak convergence result for $\sqrt{n}(\hat{\Lambda}(\cdot) - \Lambda(\cdot), \hat{\beta} - \beta_0)$. We first deal with $\hat{\xi}_p(x)$ and toward the end of the proof we switch to $\hat{\xi}^c_p(x)$. Let

$$\hat{F}(t|x) = 1 - \prod_{u\le t}(1-\hat{\Lambda}(du))^{\exp(\hat{\beta}'x)}, \qquad F(t|x) = 1 - \prod_{u\le t}(1-\Lambda(du))^{\exp(\beta_0'x)}$$

and (A.1)

$$\hat{F}(t) = 1 - \prod_{u\le t}(1-\hat{\Lambda}(du)), \qquad F(t) = 1 - \prod_{u\le t}(1-\Lambda(du))$$

In this notation, we have

$$\sqrt{n}(\hat{\xi}_p(x) - \xi_p(x)) = \sqrt{n}\Big(\hat{F}^{-1}\Big(1-(1-p)^{\exp(-\hat{\beta}'x)}\Big) - F^{-1}\Big(1-(1-p)^{\exp(-\beta_0'x)}\Big)\Big).$$

Write the identity

$$\sqrt{n}(\hat{\xi}_p(x) - \xi_p(x)) = U_n(p,x) + T_n(p,x), \tag{A.2}$$

where

$$U_n(p,x) = \sqrt{n}\Big(\hat{F}^{-1}\Big(1-(1-p)^{\exp(-\hat{\beta}'x)}\Big) - F^{-1}\Big(1-(1-p)^{\exp(-\hat{\beta}'x)}\Big)\Big)$$

and (A.3)

$$T_n(p,x) = \sqrt{n}\Big(F^{-1}\Big(1-(1-p)^{\exp(-\hat{\beta}'x)}\Big) - F^{-1}\Big(1-(1-p)^{\exp(-\beta_0'x)}\Big)\Big).$$

Throughout this proof, (p, x) ranges over $[0, p^{(1)}] \times K$. Theorems 3.2 and 3.4 of AG state that in $D[0, \tau + \epsilon] \times \mathcal{R}^q$, as $n \to \infty$

$$\sqrt{n}(\hat{\Lambda}(t) - \Lambda(t), \hat{\beta} - \beta_0) \xrightarrow{d} \Big(W(a(t)) - b(t)'\Sigma^{-1/2}Z, \Sigma^{-1/2}Z\Big), \qquad \text{(A.4)}$$

where $a(\cdot), b(\cdot)$, and Σ are defined by (2.9) and (2.8), and $D[0, \tau + \epsilon]$ is the standard Skorohod space.

Consider first $U_n(p, x)$. By (A.4) we have (via (A.1))

$$\sqrt{n}(\hat{F}(t) - F(t), \hat{\beta} - \beta_0) \xrightarrow{d} \Big(S(t)\Big(W(a(t)) - b(t)'\Sigma^{-1/2}Z\Big), \Sigma^{-1/2}Z\Big) \qquad \text{(A.5)}$$

This follows from the easy fact that

$$\sup_{0 \le t \le \tau + \epsilon} \left|\sqrt{n}\left(\hat{S}(t) - S(t)\right) - \sqrt{n}\left(\exp(-\hat{\Lambda}(t)) - \exp(-\Lambda(t))\right)\right| \xrightarrow{pr.} 0 \qquad \text{(A.6)}$$

($\hat{S}$ and S are the survival functions corresponding to $\hat{\Lambda}$ and Λ, respectively), or somewhat more elegantly, by the compact differentiability of the product integral (Gill and Johansen (1987)): (A.5) follows from (A.4) by the delta method. To deal with $U_n(p, x)$ we will combine (A.5) with the following result.

Proposition A.1 (Doss and Gill, 1989) *Let ζ be a function defined on $[0, 1]$ that has a derivative ζ' which is positive and continuous. Let $\zeta_n, n = 1, 2, \ldots$ be nondecreasing right continuous processes on $[0, 1]$ satisfying $\zeta_n(0) = \zeta(0)$ a.s. and $\sqrt{n}(\zeta_n - \zeta) \xrightarrow{d} K$ in $D[0, 1]$, where the process K has a.s. continuous sample paths. Then, for every $\epsilon > 0$*

$$\sup_{\zeta(0) \le t \le \zeta(1) - \epsilon} \left|\sqrt{n}(\zeta_n^{-1}(t) - \zeta^{-1}(t)) + \sqrt{n}\left(\frac{\zeta_n(\zeta^{-1}(t)) - t}{\zeta'(\zeta^{-1}(t))}\right)\right| \xrightarrow{pr.} 0 \qquad \text{(A.7)}$$

In particular,

$$\sqrt{n}(\zeta_n^{-1} - \zeta^{-1}) \xrightarrow{d} \frac{K(\zeta^{-1})}{\zeta'(\zeta^{-1})} \quad \text{in} \quad D[\zeta(0), \zeta(1) - \epsilon].$$

We now combine (A.7) of the proposition and (A.5), and obtain that

$$\sqrt{n}(\hat{\xi}_p(x) - \xi_p(x)) = \frac{\tilde{Q}_n(p, x)}{F'\Big(F^{-1}\Big(1 - (1 - p)^{\exp(-\beta' x)}\Big)\Big)} + R_n^{(1)}(p, x) + T_n(p, x) \qquad \text{(A.8)}$$

where

$$\tilde{Q}_n(p, x) = -\sqrt{n}\Big(\hat{F}\Big(F^{-1}\Big(1 - (1 - p)^{\exp(-\hat{\beta}' x)}\Big)\Big) - \Big(1 - (1 - p)^{\exp(-\hat{\beta}' x)}\Big)\Big) \qquad \text{(A.9)}$$

and

$$\sup_{p, x} R_n^{(1)}(p, x) \xrightarrow{pr.} 0 \qquad \text{(A.10)}$$

Let

$$Q_n(p,x) = -\sqrt{n}\Big(\hat{F}\Big(F^{-1}\Big(1-(1-p)^{\exp(-\beta_0' x)}\Big)\Big) - \Big(1-(1-p)^{\exp(-\beta_0' x)}\Big)\Big). \quad \text{(A.11)}$$

We now apply the mean value theorem to the term $T_n(p,x)$ in (A.8) and obtain

$$\sqrt{n}(\hat{\xi}_p(x) - \xi_p(x)) = \frac{Q_n(p,x)}{F'\Big(F^{-1}\Big(1-(1-p)^{\exp(-\hat{\beta}' x)}\Big)\Big)} + R_n^{(1)}(p,x) + \quad \text{(A.12)}$$

$$\left(\frac{x\pi\exp(-\beta_0' x)}{\lambda(\xi_p(x))}\right)'\sqrt{n}(\hat{\beta}-\beta_0) + R_n^{(2)}(p,x) + \frac{\tilde{Q}_n(p,x) - Q_n(p,x)}{F'\Big(F^{-1}\Big(1-(1-p)^{\exp(-\hat{\beta}' x)}\Big)\Big)}$$

where

$$\sup_{p,x} R_n^{(2)}(p,x) \xrightarrow{pr.} 0. \quad \text{(A.13)}$$

Next, we proceed to show that

$$\frac{\tilde{Q}_n(p,x) - Q_n(p,x)}{F'\Big(F^{-1}\Big(1-(1-p)^{\exp(-\hat{\beta}' x)}\Big)\Big)} \xrightarrow{pr.} 0 \quad \text{(A.14)}$$

Using the Skorohod construction we may assume without loss of generality that in (A.5) the convergence is almost sure in sup norm (since W is continuous). The continuity of $W(a(\cdot))$ implies that if $\{\eta_n\}$ is a sequence such that $\eta_n \to 0$ a.s., then $\sup_{0\le t\le \tau+\epsilon} |W(a(t+\eta_n)) - W(a(t))| \to 0$ a.s., and this gives the convergence in probability in (A.14).

Combining (A.12) with (A.10), (A.13), and (A.14) we obtain

$$\sqrt{n}(\hat{\xi}_p(x) - \xi_p(x)) = \frac{Q_n(p,x)}{F'\big(\xi_p(x)\big)} + \left(\frac{x\pi\exp(-\beta_0' x)}{\lambda(\xi_p(x))}\right)'\sqrt{n}(\hat{\beta}-\beta_0) + R_n^{(3)}(p,x) \quad \text{(A.15)}$$

where

$$\sup_{p,x} R_n^{(3)}(p,x) \xrightarrow{pr.} 0. \quad \text{(A.16)}$$

(In the denominator of the first term on the right side of (A.12), the change resulting from substituting β_0 for $\hat{\beta}$ is absorbed into $R_n^{(3)}(p,x)$).

This shows that the finite dimensional distributions of $\sqrt{n}(\hat{\xi}_p(x) - \xi_p(x))$ converge to those of

$$-\left(\frac{W\big(a(\xi_p(x))\big) - \big(b(\xi_p(x))\big)'\Sigma^{-1/2}Z}{\lambda(\xi_p(x))}\right) + \frac{\big(\pi x\exp(-\beta_0' x)\big)'}{\lambda(\xi_p(x))}\Sigma^{-1/2}Z$$

and since $-W \stackrel{d}{=} W$, this means that

> the finite dimensional distributions of $\sqrt{n}(\hat{\xi}_p(x) - \xi_p(x))$
> converge to those of $V(p,x)$. (A.17)

We now consider tightness. Let $w_v(\cdot)$ denote the continuity modulus of the function v. The weak convergence in D to a continuous process of $\sqrt{n}(\hat{S} - S)$ implied by (A.5) gives the following.

$$\text{For every } \epsilon > 0 \quad \lim_{\delta \to 0} \limsup_{n \to \infty} P\left\{ w_{\sqrt{n}(\hat{S}-S)}(\delta) \geq \epsilon \right\} = 0.$$

Therefore, if $\tilde{w}_{\cdot}(\cdot)$ denotes the continuity modulus of a function defined on the $(q+1)$-dimensional cube, (A.15) and (A.16) give

$$\text{for every } \epsilon > 0 \quad \lim_{\delta \to 0} \limsup_{n \to \infty} P\left\{ \tilde{w}_{\sqrt{n}(\hat{\xi}_p(x) - \xi_p(x))}(\delta) \geq \epsilon \right\} = 0. \tag{A.18}$$

We now switch to $\hat{\xi}_p^c(x)$, which is what the theorem actually refers to. Note that $\sup_{p,x} \sqrt{n}(\hat{\xi}_p^c(x) - \hat{\xi}_p(x)) \xrightarrow{pr.} 0$ (by (A.7) of Proposition A.1 for example), and so (A.17) is still true if we replace $\hat{\xi}_p(x)$ by $\hat{\xi}_p^c(x)$. Moreover, (A.18) implies that

$$\text{for every } \epsilon > 0 \quad \lim_{\delta \to 0} \limsup_{n \to \infty} P\left\{ \tilde{w}_{\sqrt{n}(\hat{\xi}_p^c(x) - \xi_p(x))}(\delta) \geq \epsilon \right\} = 0. \tag{A.19}$$

Now, convergence of finite-dimensional distributions together with (A.19) are enough to give the desired weak convergence.
Note: We have made use of the fact that the characterization of weak convergence in C_{q+1} is the same as the characterization of weak convergence in C given in Billingsley (1968, pp. 54–55). This is because the characterization of the compact sets given by the Arzela-Ascoli theorem is the same for the two spaces C and C_{q+1}.

Proof of Lemma 2.1

Parts (ii) and (iii) are Corollaries 3.5 and 3.3, respectively, of AG. The proof of Part (i) is very similar (cf. the second part of the proof of Theorem 3.2 of AG). Part (iv) follows directly from Theorem 1. Part (v) follows from the results of Ramlau-Hansen (1983); see in particular his Theorem 4.1.2 and also lines 8–5 from the bottom on page 460. (Actually, Ramlau-Hansen obtains uniform consistency of kernel estimates for Aalen's multiplicative intensity model. However, the extension to the Cox model does not present a serious difficulty.) Part (vi) follows from Parts (i)–(v). The proof of Part (vii) is straightforward.

Proof of Theorem 2

Theorem 2 follows directly from Theorem 1 and Part (vii) of Lemma 2.1.

Proof of Theorem 3

Let $p^{(0)}$ be fixed. We will first show that as $n \to \infty$

$$\hat{L}(p, x) \xrightarrow{d} L(p, x) \quad \text{in} \quad C([p^{(0)}, p^{(1)}] \times K). \tag{A.20}$$

For $f \in C([p^{(0)}, p^{(1)}] \times K)$ we shall denote $\sup_{p \in [p^{(0)}, p^{(1)}], x \in K} |f(p, x)|$ by $||f||$. To obtain (A.20) assume without loss of generality that $\hat{L}$ and L are defined on a common probability space. More precisely, let $(\Omega, \mathcal{F}, P)$ be the probability space on which the random vectors (Y_i, X_i, C_i) $i = 1, \ldots, n$ are defined, let $(\Omega', \mathcal{F}', P')$ be a probability space on which $W(\cdot)$ and Z are defined, and consider the product space $(\Omega \times \Omega', \mathcal{F} \times \mathcal{F}', P \times P')$. On this probability space we may define (versions of) L and $\hat{L}$ by (2.13) and (2.15), respectively. (The same $W(\cdot)$ and Z are used in the definition of L and $\hat{L}$.) Note that $||\hat{L} - L|| \xrightarrow{pr.} 0$. This follows from the continuity of $W(\cdot)$ and Lemma 2.1. This gives (A.20) and therefore that

$$||\hat{L}|| \xrightarrow{d} ||L|| \tag{A.21}$$

Letting $L_n(p, x) = \sqrt{n}(\hat{\xi}_p(x) - \xi_p(x))/\hat{\sigma}(p, x)$, Theorem 2 implies that

$$||L_n|| \xrightarrow{d} ||L|| \tag{A.22}$$

By Theorem 1 of Tsirel'son (1975) the distribution of $||\hat{L}||$ (and also that of $||L||$) is continuous. Therefore, by (A.21) and (A.22)

$$\sup_{-\infty < t < \infty} \Big| P\{||L_n|| \le t\} - P\{||\hat{L}|| \le t\} \Big| \longrightarrow 0 \tag{A.23}$$

Recalling that $s_\alpha^{(n)}$ is a $(1 - \alpha)^{th}$ quantile of $||\hat{L}||$, we see that (A.23) implies that $P\{||L_n|| \le s_\alpha^{(n)}\} \longrightarrow 1 - \alpha$, and this proves the theorem.
Note: $s_\alpha^{(n)}$ is approximated through a simulation. However, it is easy to see that this does not cause any problem in the theory.

Proof of Theorem 4

To prove Theorem 4 we shall first prove Theorem A.1, which states that as $n \to \infty$, the distribution of $\sqrt{n}(\hat{\xi}_p^*(x) - \hat{\xi}_p(x))$ converges to the limiting distribution of $\sqrt{n}(\hat{\xi}_p(x) - \xi_p(x))$. We express this notion of "weak convergence in probability" in terms of the Prohorov metric d on the space of all probability measures on C_{q+1}. This metric is defined in Billingsley (1968, pp. 236–238). The feature of it that we will use is that it metrizes weak convergence: If μ_n and μ are probability measures on C_{q+1}, then $d(\mu_n, \mu) \to 0 \iff \mu_n \xrightarrow{d} \mu$.

We will use the following notation. If $\zeta(p, x)$ is a process in C_{q+1}, then $\mathcal{L}(\zeta(p, x))$ and $\mathcal{L}(\zeta(p, x)|\text{data})$ denote the distribution of $\zeta(p, x)$ and the conditional distribution of $\zeta(p, x)$ given the data, respectively. Here, of course, the data is (T_i, δ_i, X_i), $i = 1, \ldots, n$.

Theorem A.1 *Assume the conditions of Theorem 1. Then*

$$d\Big(\mathcal{L}\Big(\sqrt{n}(\hat{\xi}_p^{*c}(x) - \hat{\xi}_p^c(x))|\text{data}\Big), \mathcal{L}\Big(V(p, x)\Big)\Big) \xrightarrow{pr.} 0 \tag{A.24}$$

where $V(p, x)$ is the process defined in Theorem 1.

The proof of Theorem A.1 is similar to that of Theorem 1. The main changes are the following.

1 We use Hjort's (1985a) result on the consistency of the bootstrap in the Cox model, instead of AG's weak convergence result (A.4).
2 We use a version of Proposition A.1 appropriate for the bootstrap.
3 We use elementary bounds on the size of the largest jump of the bootstrap version of $\hat{S}$.

Instead of indicating where changes need to be made, we give the entire proof for the sake of completeness.

Recall that $\hat{\Lambda}^*$ and $\hat{\beta}^*$ are the same as $\hat{\Lambda}$ and $\hat{\beta}$ except that they are calculated from the bootstrap sample. Define

$$\hat{F}^*(t|x) = 1 - \prod_{u \le t}(1 - \hat{\Lambda}^*(du))^{\exp(\hat{\beta}^{*\prime}x)} \quad \text{and} \quad \hat{F}^*(t) = 1 - \prod_{u \le t}(1 - \hat{\Lambda}^*(du)) \tag{A.25}$$

We have

$$\sqrt{n}(\hat{\xi}_p^*(x) - \hat{\xi}_p(x)) = \sqrt{n}\Big(\hat{F}^{*-1}\big(1 - (1 - p)^{\exp(-\hat{\beta}^{*\prime}x)}\big) - \hat{F}^{-1}\big(1 - (1 - p)^{\exp(-\hat{\beta}'x)}\big)\Big).$$

As in the proof of Theorem 1, we will use the decomposition

$$\sqrt{n}(\hat{\xi}_p^*(x) - \hat{\xi}_p(x)) = U_n^*(p, x) + T_n^*(p, x), \tag{A.26}$$

where

$$\begin{aligned} U_n^*(p, x) &= \sqrt{n}\Big(\hat{F}^{*-1}\big(1 - (1 - p)^{\exp(-\hat{\beta}^{*\prime}x)}\big) - \hat{F}^{-1}\big(1 - (1 - p)^{\exp(-\hat{\beta}^{*\prime}x)}\big)\Big) \\ &\text{and} \\ T_n^*(p, x) &= \sqrt{n}\Big(\hat{F}^{-1}\big(1 - (1 - p)^{\exp(-\hat{\beta}^{*\prime}x)}\big) - \hat{F}^{-1}\big(1 - (1 - p)^{\exp(-\hat{\beta}'x)}\big)\Big). \end{aligned} \tag{A.27}$$

The proposition of Section 3 of Hjort (1985a) states that under the assumption that the triples (T_i, δ_i, X_i), $i = 1, 2, \ldots$ are iid,

$$\mathcal{L}\big(\sqrt{n}(\hat{\beta}^* - \hat{\beta})|\text{data}\big) \xrightarrow{d} \mathcal{L}\big(\Sigma^{-1/2}Z\big) \qquad \text{a.s.} \tag{A.28}$$

Note that the quantity to the left of the arrow in (A.28) is a function of the data, and the expression "a.s." refers to the sequence (T_i, δ_i, X_i), $i = 1, 2, \ldots$. Define $W^\dagger$ and $Z^\dagger$ by

$$W^\dagger(t) = W(a(t)) - b(t)'\Sigma^{-1/2}Z \qquad \text{and} \qquad Z^\dagger = \Sigma^{-1/2}Z. \tag{A.29}$$

A careful look at Hjort's proof reveals the following.

1 Statement (A.28) can be strengthened to

$$\mathcal{L}\big(\sqrt{n}(\hat{\Lambda}^* - \hat{\Lambda}, \hat{\beta}^* - \hat{\beta})|\text{data}\big) \xrightarrow{d} \mathcal{L}\big(W^\dagger, Z^\dagger\big) \quad \text{a.s.}$$

(cf. lines 10–12 from the bottom on page 5 of Hjort, 1985a), which is equivalent to

$$d\Big(\mathcal{L}\big(\sqrt{n}(\hat{\Lambda}^* - \hat{\Lambda}, \hat{\beta}^* - \hat{\beta})|\text{data}\big), \mathcal{L}\big(W^\dagger, Z^\dagger\big)\Big) \longrightarrow 0 \quad \text{a.s.}, \tag{A.30}$$

where d is the Prohorov metric on the space of all probability measures on $C[0, \tau + \epsilon] \times \mathcal{R}^{q+1}$.

2 If the iid assumption is removed, then (A.30) is weakened to

$$d\Big(\mathcal{L}\big(\sqrt{n}(\hat{\Lambda}^* - \hat{\Lambda}, \hat{\beta}^* - \hat{\beta})|\text{data}\big), \mathcal{L}\big(W^\dagger, Z^\dagger\big)\Big) \xrightarrow{pr.} 0. \tag{A.31}$$

which is obtained by working with subsequences.
Let $\{n_k\}$ be any subsequence of $\{n\}$. To show (A.24) it suffices to show that $\{n_k\}$ has a further subsequence along which the convergence statement (A.24) holds a.s. By AG's weak convergence result (A.4) and by Hjort's result (A.31), there exists a subsequence of $\{n_k\}$ along which, with probability one,

$$||\hat{\Lambda} - \Lambda|| \longrightarrow 0 \qquad \hat{\beta} \longrightarrow \beta \tag{A.32}$$

and

$$d\Big(\mathcal{L}\big(\sqrt{n}(\hat{\Lambda}^* - \hat{\Lambda}, \hat{\beta}^* - \hat{\beta})|\text{data}\big), \mathcal{L}\big(W^\dagger, Z^\dagger\big)\Big) \longrightarrow 0 \tag{A.33}$$

($||f|| = \sup |f|$, where the set over which the sup is taken is obtained from context). In order to obtain the bootstrap analogue of (A.5), we need the following lemma. We use the notation $\Delta f(t) = f(t) - f(t-)$ for any right-continuous function with left limits.

Lemma A.1 *Under the assumptions of Theorem 4*

1 $||n^{3/4}\Delta\hat{\Lambda}^*|| \longrightarrow 0$ *in bootstrap probability, a.s.*

2 $||n^{3/4}\Delta\hat{S}^*|| \longrightarrow 0$ *in bootstrap probability, a.s.*

Lemma A.1 follows easily from Lemma A.2 below.

Lemma A.2 *Let $X_1, X_2, \ldots$ be iid from a continuous distribution F, and let F_n be the empirical distribution function of $X_1, \ldots, X_n$. Let $X_1^*, \ldots, X_n^*$ be an iid sample from F_n, and let F_n^* be the empirical distribution function of $X_1^*, \ldots, X_n^*$. Then $||n^{3/4}\Delta F_n^*|| \longrightarrow 0$ in bootstrap probability, a.s.*

Proof. Let $\epsilon > 0$. We have

$$P\{||n^{3/4}\Delta\hat{F}_n^*|| \geq \epsilon\} \leq nP\{n\Delta\hat{F}_n^*(X_1) \geq n^{1/4}\epsilon\} \leq 2n\exp(-n^{1/4}\epsilon/3),$$

the last inequality being a consequence of Bernstein's inequality (see, e.g. page 95 of Serfling, 1980).

Consider now $\hat{S}^* - \exp(-\hat{\Lambda}^*)$. We have

$$\begin{aligned}\sqrt{n}\big(\hat{S}^*(t) - \exp(-\hat{\Lambda}^*(t))\big) &= \sqrt{n}\left\{\prod_{u\leq t}(1 - d\hat{\Lambda}^*(u)) - \exp(-\Sigma_{u\leq t}d\hat{\Lambda}^*(u))\right\} \\ &\leq \sqrt{n}\big(\Sigma_{u\leq t}(d\hat{\Lambda}^*(u))^2\big),\end{aligned} \tag{A.34}$$

the inequality in (A.34) following from LeCam (1960). By Part 2 of Lemma A.1, the last expression in (A.34) converges to 0 in bootstrap probability. Combining, we see that there exists a set Ω_0 of probability one, and a subsequence $\{n_{k_j}\}$ such that

$$d\Big(\mathcal{L}\big(\sqrt{n}(\hat{F}^* - \hat{F}, \hat{\beta}^* - \hat{\beta})|\text{data}\big), \mathcal{L}\big(SW^\dagger, Z^\dagger\big)\Big) \longrightarrow 0, \quad ||\hat{\Lambda} - \Lambda|| \longrightarrow 0,$$
$$\text{and} \quad \hat{\beta} \longrightarrow \beta \quad \text{along } \{n_{k_j}\} \text{ for all data points in } \Omega_0. \tag{A.35}$$

Theorem 2 of Doss and Gill (1989) gives a version of Proposition A.1 that is appropriate for the bootstrap. Applied to our situation, it implies that there exists a subset Ω_1 of Ω_0, also of probability one, and a subsequence $\{n_{k_{j_i}}\}$ (which we will denote $\{m\}$), such that along $\{m\}$, for each data point in Ω_1

$$\Big|\Big|\sqrt{n}(\hat{F}^{*-1} - \hat{F}^{-1}) - \sqrt{n}\Big(\frac{\hat{F}^*(F^{-1}) - \hat{F}(F^{-1})}{F'(F^{-1})}\Big)\Big|\Big| \longrightarrow 0 \tag{A.36}$$

in bootstrap probability. Therefore, along $\{m\}$, for each data point in Ω_1, analogous to (A.8) we have

$$\sqrt{n}(\hat{\xi}_p^*(x) - \hat{\xi}_p(x)) = Q_n^*(p,x) + R_n^*(p,x) + T_n^*(p,x) \tag{A.37}$$

where

$$Q_n^*(p,x) = -\sqrt{n}\bigg(\hat{F}^*\Big(\hat{F}^{-1}\big(1-(1-p)^{\exp(-\hat{\beta}'x)}\big)\Big) - \big(1-(1-p)^{\exp(-\hat{\beta}'x)}\big)\bigg).$$

and

$$\sup_{p,x}|R_n^*(p,x)| \longrightarrow 0 \quad \text{in bootstrap probability.}$$

To deal with the term $T_n^*(p,x)$ we need the following fact. If c_n is an arbitrary sequence of positive constants tending to 0 then

$$\sup_{|t-u|<c_n}\Big|\sqrt{n}(\hat{F}^{-1}(t) - \hat{F}^{-1}(u)) - \sqrt{n}(F^{-1}(t) - F^{-1}(u))\Big| \xrightarrow{pr.} 0 \tag{A.38}$$

which is proved through a standard Skorohod construction argument using the continuity of the process to which $\sqrt{n}(\hat{F}^{-1} - F^{-1})$ converges. Therefore, there exists a set of probability one and a subsequence (which we shall also call $\{m\}$ and Ω_1, respectively) such that (A.38) holds along $\{m\}$, for all data points in Ω_1. This shows that T_n^* can be handled in the same way that T_n was handled, and we conclude that along $\{m\}$, for all data points in Ω_1, the finite dimensional distributions of $\sqrt{n}(\hat{\xi}_p^*(x) - \hat{\xi}_p(x))$ converge to those of $V(p,x)$.

For the rest of the proof we continue to work with the subsequence $\{m\}$ and fixed data point in Ω_1. The remaining arguments are essentially identical to those of Theorem 1, the only difference being in proving that $\sqrt{n}(\hat{\xi}_p^{*c}(x) - \hat{\xi}_p^*(x)) \longrightarrow 0$ in bootstrap probability along $\{m\}$ for all data points in Ω_1. This fact is a consequence of (A.36) and of Part 2 of Lemma A.1.

Proof of Theorem 4

Theorem 1 of Tsirel'son (1975) implies that the distribution of $||V(\cdot,\cdot)/\sigma(\cdot,\cdot)||$ is continuous. Therefore,

$$P\left\{||\frac{\hat{\xi}.(\cdot) - \xi.(\cdot)}{\hat{\sigma}(\cdot,\cdot)}|| \leq t\right\} \longrightarrow P\left\{||\frac{V(\cdot,\cdot)}{\sigma(\cdot,\cdot)}|| \leq t\right\} \tag{A.39}$$

uniformly in t. Working with uniformly convergent subsequences, we obtain a similar statement for the bootstrap analogue of (A.39). This proves the theorem.

References

Aitkin, M., Laird, N., and Francis, B. (1983), "A Reanalysis of the Stanford Heart Transplant Data," *Journal of the American Statistical Association*, 78, 264 – 291.

Andersen, P. K. and Gill, R. D. (1982), "Cox's Regression Model for Counting Processes: A Large Sample Study," *The Annals of Statistics*, 10, 1100 – 1120.

Bie, O., Borgan, O., and Liestol, K. (1987), "Confidence Intervals and Confidence Bands for the Cumulative Hazard Rate Function and Their Small Sample Properties," *The Scandinavian Journal of Statistics*, 14, 221 – 233.

Billingsley, P. (1968), *Convergence of Probability Measures*, New York: Wiley.

Breslow, N.E. (1972), "Contribution to the Discussion of the Paper by D.R. Cox," *Journal of the Royal Statistical Society*, Series B, 34, 216 – 217.

Breslow, N.E. (1974), "Covariance Analysis of Censored Survival Data," *Biometrics*, 30, 89 – 100.

Cox, D.R. (1972), "Regression Models and Life Tables" (with discussion), *Journal of the Royal Statistical Society*, Series B, 34, 187 – 220.

Cox, D. R., and Oakes, D. (1984), *Analysis of Survival Data*, London: Chapman and Hall.

Dabrowska, D. M. and Doksum, K. A. (1987), "Estimates and Confidence Intervals for Median and Mean Life in the Proportional Hazard Model," *Biometrika*, 74, 799 – 807.

Doss, H. and Gill, R. (1989), "A Method for Obtaining Weak Convergence Results for Quantile Processes, With Applications to Censored Survival Data," Technical Report M-806, Department of Statistics, Florida State University.

Efron, B. (1967), "The Two Sample Problem With Censored Data," *Proceedings of the Fifth Berkeley Symposium*, IV, 831 – 853.

— (1982), "Transformation Theory: How Normal is a Family of Distributions?" *The Annals of Statistics*, 10, 323 – 339.

Efron, B. and Tibshirani, R. (1986), "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy" (with discussion), *Statistical Science*, 1, 54 – 77.

Freedman, D. (1981), "Bootstrapping Regression Models," *The Annals of Statistics*, 9, 1218 – 1228.

Gill, R. (1984), "Understanding Cox's Regression Model: A Martingale Ap-

proach," *Journal of the American Statistical Association*. 79. 441 – 447.

Gill, R. and Johansen, S. (1987), "Product Integrals and Counting Processes," CWI Report MS-R8707, Centrum voor Wiskunde en Informatica. Amsterdam. To appear in *The Annals of Statistics*.

Hall, P. (1986a), "On the Bootstrap and Confidence Intervals," *The Annals of Statistics*, 14, 1431 – 1452.

— (1986b), "On the Number of Bootstrap Simulations Required to Construct a Confidence Interval," *The Annals of Statistics*, 14, 1453 – 1462.

Hall, W. J. and Wellner, J. A. (1980), "Confidence Bands for a Survival Curve," *Biometrika*, 67, 133 – 143.

Hjort, N. (1985a), "Bootstrapping Cox's Regression Model," Technical Report 241, Department of Statistics, Stanford University.

— (1985b), Comment on "Counting Process Models for Life History Data: A Review," by P. K. Andersen and O. Borgan, *The Scandinavian Journal of Statistics*, 12, 141 – 150.

Kalbfleisch, J. D. and Prentice, R. L. (1980), *The Statistical Analysis of Failure Time Data*, New York: Wiley.

LeCam, L. (1960), "An Approximation Theorem for the Poisson Binomial Distribution," *Pacific Journal of Mathematics*, 10. 1181 – 1197.

Marsaglia, G. and Tsang, W. W. (1984), "A Fast, Easily Implemented Method for Sampling from Decreasing or Symmetric Unimodal Density Functions," *SIAM Journal of Scientific and Statistical Computing*, 5, 349 –359.

Marsaglia, G. and Zaman. A. (1989), "Toward a Universal Random Number Generator," *Statistics and Probability Letters*. 8, 1 – 5.

Miller. R., and Halpern, J. (1982), "Regression With Censored Data," *Biometrika*. 521 – 531.

Nair, V. N. (1984), "Confidence Bands for Survival Functions With Censored Data: A Comparative Study," *Technometrics*. 26, 265 – 275.

Neuhaus, G. (1971), "On Weak Convergence of Stochastic Processes With Multidimensional Time Parameter," *The Annals of Mathematical Statistics*. 42, 1285 – 1295.

Ramlau-Hansen, H. (1983), "Smoothing Counting Process Intensities by Means of Kernel Functions," *The Annals of Statistics*. 11, 453 – 466.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

Silverman. B. W. (1986), *Density Estimation for Statistics and Data Analysis*.

London: Chapman and Hall.

Tanner, M. A. (1984), "Data-based Non-parametric Hazard Estimation," *Applied Statistics*, 33, 248 – 258.

Tsiatis, A. (1981), "A Large Sample Study of Cox's Regression Model," *The Annals of Statistics*, 9, 93 – 108.

Tsirel'son, J. L. (1975), "The Density of the Distribution of the Maximum of a Gaussian Process," *Theory of Probability and its Applications*, 20, 847 – 856.

Whittemore, A. S. and Keller, J. B.(1986), "Survival Estimation Using Splines," *Biometrics*, 42, 495 – 506.

Wu, C.F.J. (1986), "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis" (with discussion), *The Annals of Statistics*, 14, 1261 – 1350.

Table 1. A Monte Carlo study of the three bands, survival distribution Exponential(1), censoring distribution exponential (mean 2.49; 20% censoring), covariate evenly-spaced on $(0,1), n = 80$.

	% Covering (s.c.)			Mean Width (s.c.)			Mean Critical Constant (s.c.)	
$1-\alpha$	SP1	SP2	B	SP1	SP2	B	SP1	B
.95	.944(.002)	.936(.002)	.966(.003)	.527(.001)	.566(.001)	.640(.002)	2.60(.001)	3.20(.005)
.90	.902(.003)	.898(.003)	.928(.004)	.469(.001)	.496(.001)	.538(.002)	2.32(.001)	2.69(.004)

Note: SP1 denotes the untransformed equal-precision simulated-process band, SP2 denotes the log-transformed equal-precision simulated-process band, B denotes the bootstrap equal-precision band. No information on the SP2 critical constants is given because the critical constants for SP2 are the same as the critical constants for SP1. The number of simulations is 10,000 for SP1 and SP2 and 5,000 for B.

Table 2. Monte Carlo studies comparing mean integrated squared errors of the three estimators of median survival time.

Survival Dist.	% Censoring	range of x	n	$\bar{\xi}^c_{1/2}(x)$	$\tilde{\xi}^c_{1/2}(x)$	$\hat{\xi}^c_{1/2}(x)$	Obs. %Reduction in m.i.s.e. of $\hat{\xi}$ from $\bar{\xi}$
Exp(1)	10	(0,1)	20	3.93(.060)	3.69(.053)	3.71(.054)	5
			50	1.50(.019)	1.47(.019)	1.47(.019)	2
		(-1,1)	20	21.8,14.9(.429)	20.0,13.8(.446)	13.9,11.4(.207)	36,23
			50	8.19(.125)	7.81(.116)	7.03(.096)	14
	20	(0,1)	20	5.18(.188)	4.59(.119)	4.62(.123)	11
			50	1.62(.022)	1.58(.020)	1.59(.021)	2
		(-1,1)	20	23.2,14.5(.708)	25.7,13.8(2.27)	15.8,11.4(.951)	32,21
			50	8.89(.138)	8.43(.129)	7.35(.097)	17
Weibull	10	(0,1)	20	.71(.008)	.69(.008)	.69(.008)	3
(1.35,2.0)			50	.30(.004)	.29(.003)	.29(.003)	7
		(-1,1)	20	2.01(.024)	1.94(.025)	1.67(.017)	17
			50	.96(.012)	.93(.012)	.87(.010)	10
	20	(0,1)	20	.77(.009)	.74(.008)	.74(.008)	4
			50	.31(.003)	.31(.003)	.31(.003)	1
		(-1,1)	20	2.58,1.94(.079)	2.54,1.85(.083)	1.85,1.57(.024)	28,19
			50	.94(.011)	.92(.011)	.85(.009)	9

Note: The table entries are 100×m.i.s.e.(100×std. error of m.i.s.e.). In rows 3, 7, and 15, the second numbers are the 5% trimmed means of integrated squared error. The number of simulations in each of these studies is 10,000.

Table 3. 95% bands and intervals in SHT data, for median survival time in years.

Age (Years)		Band	Interval
38.5	SP1	(1.6, 9.7)	(2.4, 6.7)
	SP2	(1.7, 10.2)	(2.4, 6.9)
	B	(1.8, 9.1)	(2.3, 5.6)
48.7	SP1	(.5, 4.8)	(.8, 2.8)
	SP2	(.5, 5.3)	(.8, 2.9)
	B	(.3, 6.4)	(.6, 2.2)

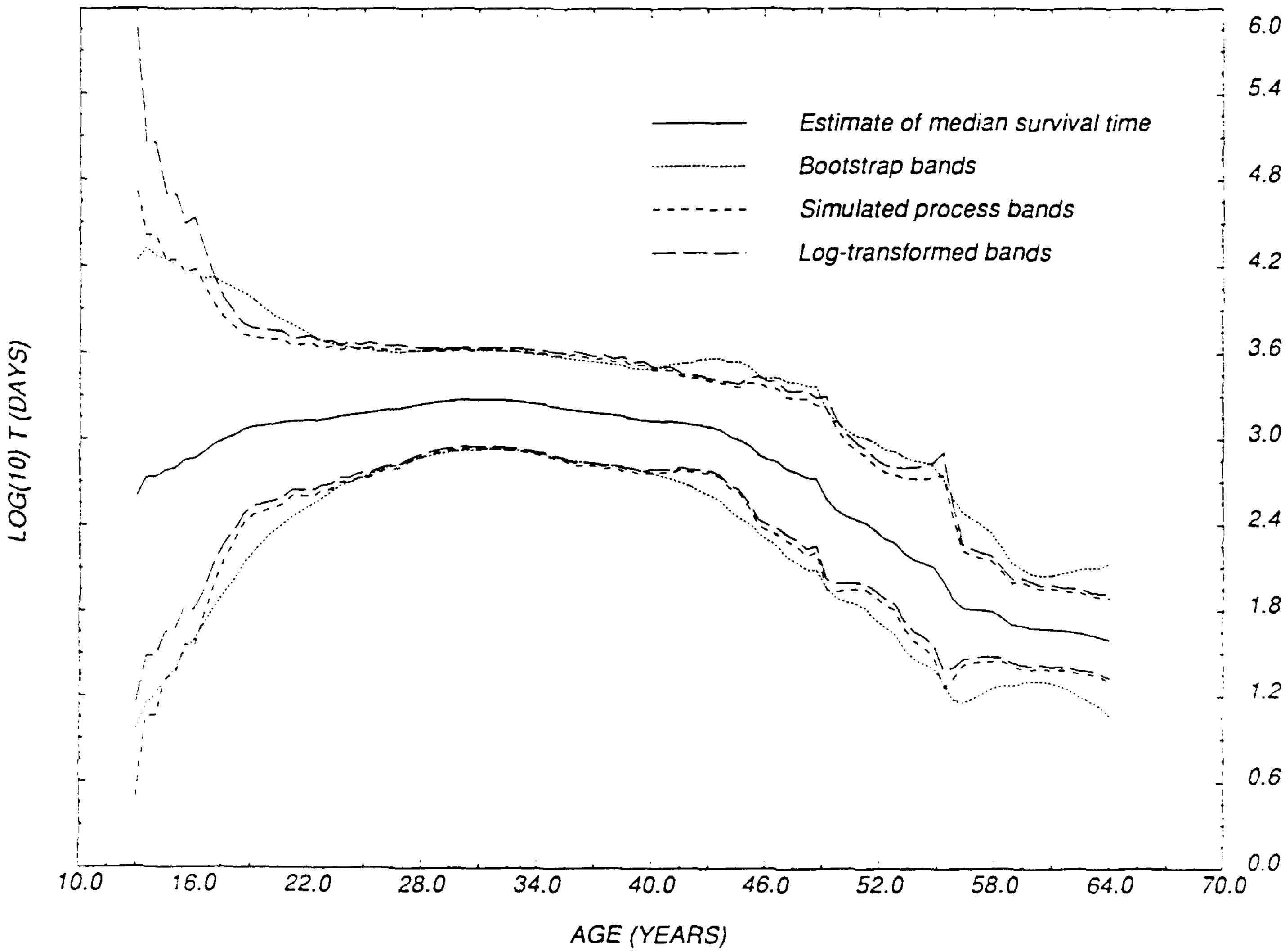

Figure 1. 95% confidence bands for median survival time in SHT data

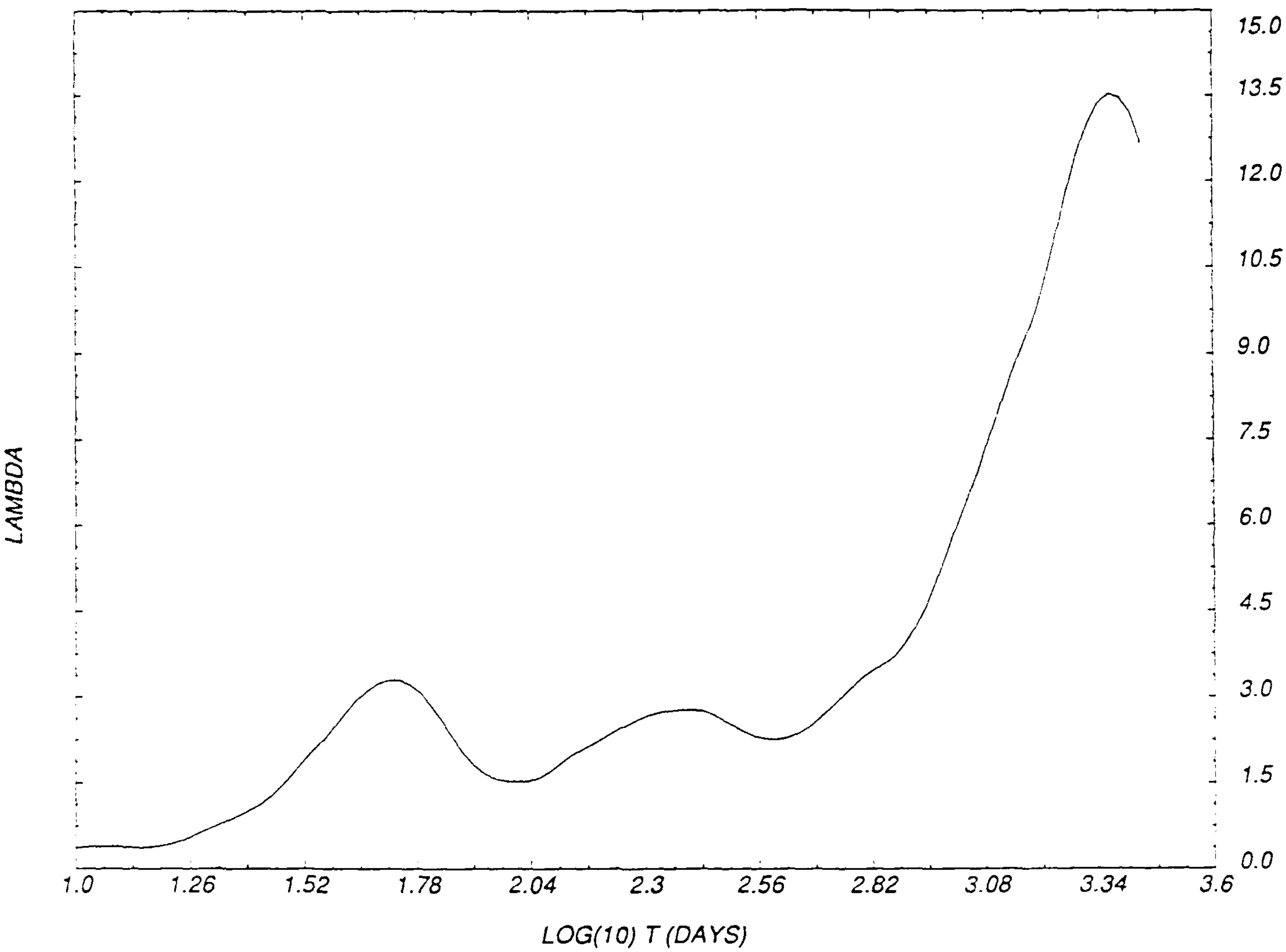

Figure 2. Kernel estimate of the hazard rate, SHT data

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 443	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Confidence Bands For The Median Survival Time As A Function Of The Covariates In The Cox Model		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Deborah Burr and Hani Doss		8. CONTRACT OR GRANT NUMBER(s) N00014-89-J-1627
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 1111		12. REPORT DATE June 25, 1991
		13. NUMBER OF PAGES 38
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Also issued by Florida State University with support of AFOSR Grant No. 90-0202.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Gaussian process simulation, bootstrap confidence bands, quantiles.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

PLEASE SEE FOLLOWING PAGE.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TECHNICAL REPORT NO. 443

20. ABSTRACT

Let $\xi_p(x)$ be the p^{th} quantile of the distribution of the lifelength of an individual with covariate vector x in the Cox model. We introduce an estimator $\hat{\xi}_p(x)$ of $\xi_p(x)$ and develop several families of confidence bands for $\xi_p(x)$ as a function of x. To construct one type of band we proceed as follows. We show that as $n \to \infty$, where n is the number of individuals in the study, $\sqrt{n}(\hat{\xi}_p(x)-\xi_p(x))$ converges weakly to a Gaussian process $W(x)$ with a complicated covariance structure. We then estimate this covariance structure from the data, and simulate many Gaussian processes with this estimated covariance structure. The critical constants required for the construction of the confidence bands are obtained from the simulated processes. Another type of bands is obtained by bootstrapping. We obtain an asymptotic theory for both types of bands. Simulation studies are used to compare the two types of bands. The methods are illustrated on the Stanford Heart Transplant Data.